# The Virtual Caliper: Rapid Creation of Metrically Accurate Avatars from 3D Measurements

Sergi Pujades, Betty Mohler, Anne Thaler, Joachim Tesch, Naureen Mahmood, Nikolas Hesse, Heinrich H Bülthoff, Michael J Black

# The Virtual Caliper: Rapid Creation of Metrically Accurate Avatars from 3D Measurements

Sergi Pujades, Betty Mohler, Anne Thaler, Joachim Tesch, Naureen Mahmood, Nikolas Hesse, Heinrich H. Bülthoff, and Michael J. Black

Fig. 1. The Virtual Caliper. A novice user takes 3D measurements of her body using the wand controllers of the HTC Vive in just under 5 minutes by following simple instructional videos. Additionally, the user is asked to enter her weight and gender. The Virtual Caliper immediately produces a rigged 3D model with exactly the dimensions of the measured person. The rigged 3D model can be used straight away in Unity or exported for use in other 3D modeling packages. We illustrate a virtual mirror application where the user can look down and experience her body in VR.

**Abstract**— Creating metrically accurate avatars is important for many applications such as virtual clothing try-on, ergonomics, medicine, immersive social media, telepresence, and gaming. Creating avatars that precisely represent a particular individual is challenging however, due to the need for expensive 3D scanners, privacy issues with photographs or videos, and difficulty in making accurate tailoring measurements. We overcome these challenges by creating "The Virtual Caliper", which uses VR game controllers to make simple measurements. First, we establish what body measurements users can reliably make on their own body. We find several distance measurements to be good candidates and then verify that these are linearly related to 3D body shape as represented by the SMPL body model. The Virtual Caliper enables novice users to accurately measure themselves and create an avatar with their own body shape. We evaluate the metric accuracy relative to ground truth 3D body scan data, compare the method quantitatively to other avatar creation tools, and perform extensive perceptual studies. We also provide a software application to the community that enables novices to rapidly create avatars in fewer than five minutes. Not only is our approach more rapid than existing methods, it exports a metrically accurate 3D avatar model that is rigged and skinned.

**Index Terms**—Full body avatars, metric accuracy, rapid creation.

## 1 INTRODUCTION

Creating realistic full-body 3D avatars is important for many applications such as virtual clothing try-on [37], ergonomics [11, 33] medicine [36, 47], telepresence [24, 39] and gaming. While there has been extensive research and there are many commercial systems for facial avatar creation, the full body has received less attention. Current approaches require either skilled avatar creators (artistic approach), anthropometric measurements from trained personnel (anthropometric approach), photographs (image-based approach), or specialized hardware and software (technological approach). In all cases, current avatar creation approaches involve a time-consuming pipeline and/or require expertise. However, many applications require not only real-time creation of avatars, but also real-time performance in terms of pose and animation to realize their full potential.

In this paper, we present The Virtual Caliper, a tool that empowers novice users to create a metrically precise avatar in real-time with their own hands. We exploit an inexpensive and easy-to-use virtual reality (VR) system (HTC Vive) to make simple body measurements. We also provide a desktop tool for creating bodies and editing body shape. The surface of the resulting avatar matches the users' measured dimensions and volume. The exported body model is rigged and can be animated in real time for VR applications.

There are many pros and cons to consider when designing an avatar creation tool. One major issue, which has been disregarded by most methods, is the protection of the privacy and identity of the individual. To assess the need for privacy-preserving methods, we conducted an online survey (52 participants, 27 female) that asked what data people would be willing to share for different purposes such as buying clothes online, medical applications, etc.: 80.8% preferred to share metric body measurements, followed by photographs in clothing (17.3%). People least preferred (94.2%) to share videos of themselves in underwear, which is required today for the methods that generate the most metrically accurate avatars from 3D scanning systems. Currently no tool exists with which novice users can create reliably precise avatars while protecting privacy and identity.

Additional factors need to be considered by any practical solution for


- *Sergi Pujades is with Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, Grenoble, France. The work was done at Max Planck Institute for Intelligent Systems. E-mail: sergi.pujades-rocamora@inria.fr.*
- *Betty Mohler is with Amazon. E-mail: btmohler@amazon.com*
- *Anne Thaler is with Max Planck Institute for Biological Cybernetics. E-mail: anne.thaler@tuebingen.mpg.de*
- *Joachim Tesch is with Max Planck Institute for Intelligent Systems. E-mail: joachim.tesch@tuebingen.mpg.de*
- *Naureen Mahmood is with Meshcapade. E-mail: naureen@meshcapade.com.*
- *Nikolas Hesse is with Fraunhofer Institute of Optronics, System Technologies and Image Exploitation. E-mail: nikolas.hesse@iosb.fraunhofer.de*
- *Heinrich H. Bülthoff is with Max Planck Institute for Biological Cybernetics. E-mail: heinrich.buelthoff@tuebingen.mpg.de.*
- *Michael J. Black is with Max Planck Institute for Intelligent Systems. E-mail: black@tuebingen.mpg.de.*


*Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx*

| | Artistic | Anthropometric | Image-Based | Specialised Technology | Virtual Caliper |
|---|---|---|---|---|---|
| Under 5 minutes | ✓ | ✗ | ✓ | ✓ | ✓ |
| Precise avatars | ✗ | ✓ | ✓ | ✓ | ✓ |
| Protects Privacy | ✓ | ✓ | ✗ | ✗ | ✓ |
| Freely Available | ✓ | ✗ | ✓ | ✗ | ✓ |
| Novice Users | ✓ | ✗ | ✓ | ✗ | ✓ |

Fig. 2. Pros (Green) and cons (Red) for artistic methods, methods using anthropometric measurements, image-based methods and specialized technology. The Virtual Caliper was designed to fulfill them all.

full-body avatar creation. We cannot assume user expertise in 3D modeling or artistic talent. The cost and/or availability to the public are key to adoption. The process of creating an avatar should be fun, easy, and quick. Finally, it is important that the avatar creation process is computationally lightweight and allows creators access to the virtual avatar for other applications, not just as an image or embedded in an application. The summary in Figure 2 shows the requirements to consider when choosing an avatar creation tool. Our proposed approach fulfills them all. To make the Virtual Caliper, we address a key research question: *what measurements are good for creating an avatar of one's own body?* Previous research has, quite naturally, focused on standard tailoring measurements. Several methods have shown that relating a statistical model of 3D body shape to such measurements works well [31, 59, 68]. There are, however, several problems with tailoring measurements that have not previously been elucidated. First, such measurements require expertise to be taken precisely. Even trained experts vary significantly in measuring the same person [26, 52]. Second, tailoring measurements must be taken by another person; it is not possible to accurately measure one's own body.

Here we observe that the controllers for VR applications can be repurposed to act as a measuring device. Specifically with the HTC Vive, the precise location of the wand controllers is known[1]. So, instead of assuming the standard tailoring measurements, we explore *what measurements can be reliably taken with these controllers on the own body.* Girths, or circumference measurements, are difficult, but distance measurements are not. Consequently, we selected a set of 13 distance, or point-to-point, measurements on the body and had a group of subjects perform these. From this set, we found a subset of 4-5 that can be made accurately and repeatably by novices with only short video instructions. The process is quick and easy, taking roughly 5 minutes. Figure 1 illustrates The Virtual Caliper measuring process.

The next question is *whether such distance measurements can be used to create realistic avatars.* To relate distance measurements to 3D body shape, we first need an appropriate representation of shape. By appropriate here, we mean one that is naturally related to distance measurements on the body. Models like SCAPE [10] are based on triangle deformations and thus live on a non-Euclidean manifold [23]. An Euclidean representation of shape is preferred because distance measurements are then directly relatable to the surface representation, simplifying the mapping from one to the other. We demonstrate that distances between points on the SMPL body model [44] are directly linearly related to its shape parameters. Thus we can rely on linear regression to construct body shapes from point-to-point distances. This simple regression has two major advantages: it can be easily integrated in any other software, and allows real-time performance.

To evaluate the metric accuracy of The Virtual Caliper, we compare the created bodies to ground truth 3D scan data and several other avatar creation methods. We show that our method offers the best trade-off of accuracy and ease of use.

The created avatar is rigged and can be directly animated. The joint locations are inferred from the avatars' shape using the SMPL body model [44], which was learned from the CAESAR database [54] containing around 2,000 scans per gender. The evaluation of the precision of these joint locations and the challenging problem of inverse kinematics are beyond the scope of this work. The resulting avatars are not textured, as one motivational element of this work is the protection of the users' privacy by not using any picture. In Section 6 we discuss how, with the users' consent, the obtained avatar could be textured.

In summary, our contributions include: 1) Evaluating what measurements can be made accurately on the body using the HTC Vive; 2) Defining body measurements (based on both consistency and ease of measuring) that are linearly related to 3D body shape; 3) Creating a simple protocol and software to capture these measurements; 4) Learning linear regressors to quickly produce 3D body shapes using the measurements; 5) Validating our approach metrically and perceptually in extensive user studies; and 6) Releasing these software tools to the research community, enabling anyone to create accurate 3D avatars that can be immediately animated. The desktop and HTC Vive software tools are available at `www.thevirtualcaliper.is.tue.mpg.de`.

## 2 Related Work

There are multiple ways to create a human avatar. While several approaches have focused exclusively on facial avatars [14, 21, 42], we focus our scope to full-body avatar creation methods. We present them in the following categories: artistic approaches, approaches relying on a body shape space and anthropometric measurements, approaches using images or depth sensor data, and technological approaches. Finally we review applications that require accurate avatars.

### 2.1 Artistic approaches

Many avatar creation tools are available, such as AXYZDesign [2], Mixamo [6], Poser [7] and MakeHuman [5]. Artistic approaches allow a user to enter values for the body dimensions of the avatar. Artists use their training and skilled judgment of human shape to define body proportions. But, as we will see, novices are not very good at producing bodies based on their judgement of dimensions. Another possibility is to simply measure bodies, however measuring them precisely is challenging. Artists will often instead use ratios, or average values as an initial guess.

### 2.2 Body Shape Space Based Approaches

The pioneering work of Blanz et al. [14] proposes a method to map facial attributes, such as femininity, to the shape space of their 3D face model. Since then, many methods have approached the problem of body shape space reparametrization [9, 19, 34, 57, 73]. The idea is to learn a mapping between semantic labels and the body model space. By modifying these semantic features the user can explore the human body shape space. For instance, in Body Talk [62] a plausible 3D body is created with words that describe body shape. The focus of these approaches is usually visual plausibility and ease of use. Although height and weight are usually used, the approaches mostly use semantic values that are difficult to quantify. Thus the methods may not achieve metric precision.

Other research efforts have specifically focused on the regression of avatars from body measurements by linking these measurements with a representation of the shape space. These approaches show that regressing bodies from accurate measurements is possible and works well. Seo and Magnenat-Thalmann [59] approach the creation of new bodies as a data sample interpolation problem, and their interpolants are accurate for circumference predictions. Hasler et al. [31] train semantically meaningful regressors to generate novel meshes according to several constraints and also report metric accuracy for different linear and non-linear regression techniques. Wuhrer et al. [68] learn a mapping between measurements and bodies from the CAESAR dataset [55]. As the measured bodies may lie outside their body shape space, they propose a refinement step that involves optimization, which is not well suited for real-time applications. In our work we also train precise

[1]The tracking system of the HTC Vive was created by Steam VR.

- and yet simpler - regressors, but we focus on the measurements that can be best performed by novice users.

Interesting related works on anatomical models have been proposed [35, 58]. For instance Saito et al. [58] create a personalized anatomical avatar by measuring some dimensions, such as bone lengths. The created models are targeted towards physics simulation.

Anthropometric measurements. Most of the methods that regress body shape from measurements rely on anthropometric measurements, as these remain the gold standard for body measurement. Without internal scanning of bone lengths, all dimensions are only estimates. Unavoidably, there is variability between measurers and by the same measurer over time [52]. Typical certified anthropomorphic programs, which teach how to measure the body, are costly, and time consuming. A common measurement protocol involves taking the same measurement twice. If the relative error between both measurements lies inside a certain threshold, then the mean of the measurements is taken. In other protocols (i.e. ANSUR [26] and ISAK [4]), a third measurement is taken and the median of the 3 values is used as the final value. We adopt the ISAK method for our measurements.

### 2.3 RGB and RGB-D Approaches

Several approaches use video cameras, either image based (RGB) or image and depth sensor (RGB-D), to estimate the human body shape and create avatars. Some methods rely on single or multiple images to compute the pose and shape of a body [12, 18, 27, 29]. Guan et al. [27] estimate the shape and pose of a body from a single image. To account for the ambiguity between the height and distance from the camera, the user inputs the height of the person to estimate a height-constrained body shape. Boisvert et al. [18] compute a metrically accurate reconstruction of the body shape from two silhouettes, assuming a perfect pose of the body. However the precision is highly influenced if the pose at acquisition time is not accurate. Recent methods use convolutional neural networks in order to infer the pose and shape of the human body from a single image [17, 22, 40, 53]. As Lassner et al. [40] provide open-source online code, we compare to them.

The methods using a depth sensor in addition to the images usually combine silhouette information, together with the depth and color data [66, 70, 72]. The method by Bogo et al. [16] achieves accuracy on the order of 3*mm*, but requires the user to be in minimal clothing, and is computationally expensive. Li et al. [41] provide an original solution to create 3D self-portraits with a low cost depth sensor. While the method achieves impressive accuracy for static objects, the accuracy drops when scanning a real human. Bauer et al. [13] propose an anatomic mirror. The length of the bones are estimated from an RGB-D sequence and used to register an anatomical model. Tong et al. [63] use a fixed setting of 3 Kinect sensors to rapidly create an avatar in about 6 minutes. The obtained results are visually appealing. The reported errors comparing the results to actual anthropometric measurements on the bodies are 3cm in arm length and 2cm in leg length. In our method, as we purposefully focus on the anthropometric measurements, we obtain similar errors (1.3 cm in arm length, 2cm in leg length) with a significantly simpler setup.

### 2.4 Technological Approaches

3D Body Scanning. While 3D body scanning technologies are becoming less expensive, they still involve specialized equipment, trained technicians and heavy computations on powerful computers. Several studies compare the performance of commercial 3D scanning systems relative to hand measurements [28, 56]. As of today scanning technologies achieve the best results in terms of accuracy. From early setups [32] to more recent ones [45], the technology has evolved, allowing the creation of avatars, including the face, in one single shot.

Several approaches describe a pipeline to rapidly produce an avatar from 3D body and face scan data [8, 60]. These approaches include facial and many features of the avatar and have proven to be very useful, as they allow for instance the study of the impact of the avatar personalization on virtual body ownership [65]. However, the creation of the avatar still takes about 10 minutes and the methods are only semi-automatic. In addition they all suffer from privacy issues, as in order to obtain a metrically accurate avatar the subject needs to be scanned in minimal clothing condition. Blom et al. [15] present a pipeline for generating look-alike avatars using a combination of artistic, depth, and image-based approaches; they perceptually evaluate their avatars to be 7.5 out of 10 on a likeness rating. A difficulty with photo-realistic body scans can sometimes be the Uncanny Valley, or the sensation that the avatar is strange because it approaches life-like visual quality but does not reach it [48].

As our survey shows, users prefer to be scanned in clothing rather than in tight-fitting clothes. Several works have addressed the challenging problem of computing the shape of a person from only scans of the person wearing clothes. Existing methods tackle this problem either from images [12], a single scan [30, 53], by averaging the results on each frame [67], or by considering a full sequence of scans [50, 69, 71]. Stoll et al. [61] estimate the human shape with the focus on providing plausible animations including collision detection. Zhang et al. [71] report accuracy on the order of 3*mm* by using scans in multiple poses.

Although not a scientific publication, in [1] a tutorial shows how to use basic scaling to manually adjust the height and the arm length of the avatar using the HTC Vive wands. The body is uniformly scaled in the three dimensions to match the heights of the avatar and user. The arms are individually scaled (the rest of the body is not modified) by visually matching the 3D locations of the HTC wands and the virtual avatar; no 3D measurements are taken. While the height and armspan measurements are matched, the rest of the body is not captured and the correlations between different measurements are not modeled.

### 2.5 Application areas involving avatars

One major effort of our work targets the easy distribution of the developed tools. Thus we describe relevant application areas where researchers could benefit from the proposed tools.

Ergonomics. Avatars for ergonomic assessment and product design have been an important part of the design process since Badler created "Jack" [11] in the late 90s. Today virtual humans are used in design process to assess the life-cycle of a product from production design, evaluation, and validation [33].

Clothing applications & Virtual Try-on. Kristensen et al. [37] discuss next generation on-line shopping apparel requirements and report that an estimated 15-40% of apparel purchased on-line is returned because the clothing does not not fit or look right. A virtual and metrically accurate avatar for trying-on clothes before purchase could play an important role in reducing these costly returns/exchanges.

Medical Applications. Some medical/clinical applications require precise body models to investigate the patient's body perception such as in Anorexia Nervosa [36, 47]. Some authors have used MakeHuman [5] among other avatar creation tools, to investigate self-body perception. The assessment of body image disturbance in patient populations relies upon precise tools to quantify distortions. For all applications involving patients, privacy and the protection of self-identifying information (such as images) is essential.

Telepresence. For over twenty years scientists have worked towards establishing a telepresence system that enables a person to be present digitally somewhere else in 3-dimensions both in shape and as an interactive viewer [24, 39]. Fuchs and colleagues describe the greatest challenge to be the real-time demands of telepresence (immersive virtual reality) systems [24]. The combination of model-based and video-based approaches to creating avatars may be a possible hybrid solution for real-time performance demands.

## 3 Method

In Figure 3 we present the overview of our method. First, we define and identify 3D measurements that can be easily performed on the real (Sec. 3.1) and virtual (Sec. 3.2) bodies. Then we find their relations (Sec. 3.3). Based on the identified reliable measurements we train four regressors (Sec. 3.4) which create metrically accurate personalized avatars from sparse 3D measurements. We evaluate the generated avatars both metrically and perceptually (Sec. 4).

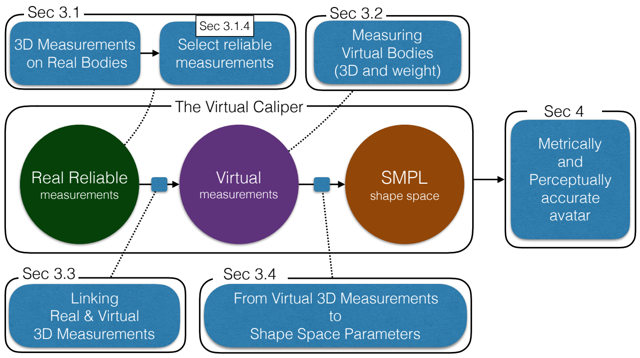


Fig. 3. The Virtual Caliper overview.

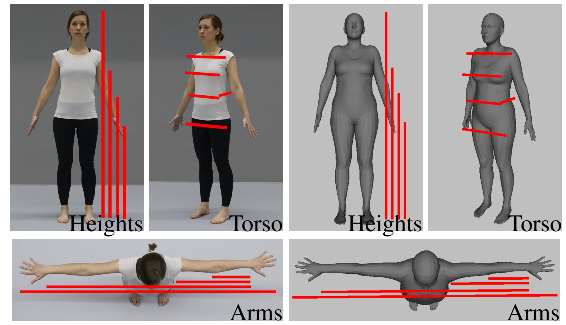


Fig. 4. The studied 3D Measurements: four heights, five in the torso and four in the arms. Left: real body. Right: virtual body.

### 3.1 3D Measurements on Real Bodies - The Virtual Caliper

Our first goal is to identify real, easy to instruct measurements on the human body that laypersons can perform on their own in a repeatable and accurate way.

#### 3.1.1 The Virtual Caliper

In our work, we propose to take advantage of the new proliferating technologies, such as the HTC Vive. We develop a tool in the Unity game engine for making rapid and precise measurements of the human body: The Virtual Caliper. The tool uses the SteamVR tracking technology of the HTC Vive system. It allows the user to make position and distance measurements on the human body using the two HTC Vive wand controllers. The measurement setup uses two SteamVR Lighthouse basestations. An initial floor calibration routine ensures accurate height measurements by guiding the participant to measure 5 different locations on the floor around the center of the tracking space. To take each measurement, the user places the tip of the wand controller handle on the floor surface and presses the trigger button for one second. We observe that this trigger hold time is important to reduce the number of false samples. Successful data capture of the wand controller tip location is confirmed with visual and auditory signals. After the floor calibration, the floor offset is calculated as the median height of the five samples and applied to the wand controller pose to measure the correct height above ground. The tool then uses instructional videos to guide the user through the measurement steps.

#### 3.1.2 Defining 3D Measurements

In total, we explore thirteen 3D measurements that can be easily measured on real as well as on virtual bodies. They are distributed in four height measurements and nine distance measurements. The measured heights are: overall height, nipple height, navel height and inseam height. Five other distances focus on the torso region: shoulder width, chest width (at nipple height), waist width (at navel height), torso depth (at navel height) and hip width (at inseam height). Four arm lengths were measured by holding the two controllers in a special way: arm span fingers (finger to finger), arm span wrist (wrist to wrist), arm length (shoulder to wrist) and forearm length (inner elbow to wrist) (see Fig. 4, left). The virtual 3D measurements are defined on the SMPL mesh template with simple vertex to vertex Euclidian distances (see Fig. 4 right). We manually identify the desired vertex indices.

#### 3.1.3 Data Acquisition

To identify which measurements can be repeatably and accurately performed on the self, we conducted a user study. We collected data from 20 participants who used our new 3D measuring device. In addition to the 3D measurements, the participants' bodies were 3D scanned, anthropometric measurements were performed, and photographs were taken in different clothing conditions (see Fig. 8), enabling the quantitative evaluation of our proposed approach. Twenty participants (gender-balanced, mean age = 29.15 (SD= 6.92), mean BMI = 21.08 (SD = 2.44), two left handed and all with normal or corrected-to-normal vision) took part in the experiment. The experimental protocol was approved by the local ethics committee and was performed in accordance with the Declaration of Helsinki. All participants gave written informed consent for their participation.

In order to study repeatability, each of the 13 body measurements was repeated three times. To study accuracy, in addition to the participant measures (self-measures), a coordinator also measured each participant (other-measures) three times. We refer to the measures of the user taken by the user themself as *self*. We refer to the measures of the user taken by the coordinator as *other*. We emphasize that the coordinator does not have any anthropometric certification.

A total of 78 measurements per participant were taken (3 repetitions $\times$ two measurers $\times$ 13 3D measurements). The self-measures took $\approx$11 sec per measurement, while the other-measures took $\approx$18 sec per measurement.

To assess *repeatability*, for each subject and 3D measurement, we compute the range of the 3 repetitions by taking the difference between the maximum and the minimum. To obtain comparable values across measurements and subjects we normalize the range, dividing by the mean of the 3 measurements. Then, for each 3D measurement, we compute the mean and standard deviations of all values (see Fig. 5 left). We further studied these deviations by the measurer condition, i.e. *self* or *other* (see Fig. 5 middle). The 3D measurements are sorted from the most repeatable to the least repeatable according to the *other* condition as this shows potentially how repeatable the measurements are with the proper instruction.

To assess *accuracy*, we start by computing the 3D measurement value using the ISAK protocol from the 3 repetitions of each 3D measurement. If the two first measurements are below 1.5% difference, their mean is computed. If they are over 1,5% difference, a third measurement is taken and the median of the three is used. We are now interested in the deviation between the *self* and *other* measurements, so we compute their relative error by considering the *other* measure as the reference (see Fig. 5 right).

#### 3.1.4 Selected 3D Measurements

We want our measurements to be representative of different body parts. *Overall Height* and *Arm Span Fingers* are the most repeatable and accurate. For the torso height we decided to retain *Inseam Height*. Although *Navel Height* and *Nipple Height* are slightly more accurate, we selected *Inseam Height*, as it is more relevant for the clothing industry [20]. We did not select *Arm Span Wrist* as *Arm Span Fingers* is already accounting for the Arm Span. The next best measurements are *Arm Length* and *Hip Width*. We decided to include them both. The rest of the measurements (*Chest Width*, *Shoulder Width*, *Torso Depth*, *Forearm Length* and *Waist Width*) were not retained as they were too inaccurately performed. The errors in waist and chest regions were expected, as even personnel trained in anthropometric measurements identify the waist and chest as the most variable dimensions [52]. In Section 6 we provide suggestions on how to improve the repeatability and accuracy of these or other measurements.

The five selected 3D measurements are *Overall Height*, *Arm Span*

*Fingers*, *Inseam Height*, *Hip Width*, and *Arm Length*. Additionally we also use the weight of the subject as it is a fairly easy measure to obtain in any household or laboratory.

### 3.2 Measuring Virtual Bodies

In order to transfer real measurements to a virtual body, we need to make the corresponding measurements on the virtual body. While computing 3D distances between two vertices is easy, measuring the weight of a virtual body is not straightforward.

To obtain an approximation of the weight of a virtual body we use the relation between the weight and the volume. We register the SMPL body model to all female and male subjects of the CAESAR dataset [55], which contains the weight measurement of all the subjects, and compute the volume of the obtained meshes (registrations). Figure 6 shows a scatter plot of the weight of the female and male CAESAR subjects as a function of the volume of the SMPL fit. A clear linear relation appears, and we learn two simple linear regressors using iterative reweighted least squares (IRLS). We learn one for the female and one for the male subjects as the correlations are slightly different for each gender. The learned regression values for females are (coef. 1001.44, intercept -2.36) and for males (coef. 1056.44, intercept -5.28). The estimated density values match existing clinical estimations [38].

### 3.3 Linking Real and Virtual 3D Measurements

So far we can perform repeatable and accurate 3D measurements on the real humans as well as on the virtual bodies. However, it is unclear whether these real and virtual measurements are consistent. To explore this consistency we aligned the SMPL body model to the scans of the 20 participants and virtually measure the SMPL meshes. For two measurements we observe significant differences: *Hip Width* and *Arm Length* (see Fig. 7). Indeed, when performing these measurements the pose of the bodies in the real world is not the same as in the virtual world. For instance, the *Hips Width* is measured in the real world by instructing the participant to *stand with both feet together*, However, the virtual meshes are measured with the feet at T-pose (see Fig. 4). Thus, we learn two linear mappings between the 3D measures obtained with the real humans and the virtual ones. One for *Hips Width* and one for *Arm Length* (see Fig. 7). For evaluation we used a K-1-fold (leave one out) split for train and test.

### 3.4 From Virtual 3D Measures to Shape Space Parameters

The last step of our method is to find the relation between the 3D measurements and the body shape space. We first review the SMPL body model, then we analyze the relation between the measurements and its shape space, and conclude with the computation of the regressors.

#### 3.4.1 The SMPL Body Model

Virtual bodies may have many representations. We are interested in these that are parametrized with a low-dimensional shape space. While several body models exist that represent human body shape in a low dimensional shape space [10, 34, 49], our choice is strongly motivated by the Euclidean architecture of the SMPL shape space, which does not contain pose variation [44].

In summary, the SMPL model is a function $M(\beta,\theta)$ that takes pose ($\theta$) and shape ($\beta$) parameters and produces a watertight triangulated mesh $\mathscr{M}$ with $N = 6890$ vertices and $F = 13,776$ triangles. The SMPL model function outputs the vertex locations of the triangulated surface. The shape parameters $\beta$ are the coefficients of a low-dimensional shape space, learned from thousands of registered scans. In this work we use 10 coefficients: $\beta \in \mathbb{R}^{10}$. The pose of the body is determined by angular rotations in a kinematic structure containing $K = 23$ joints. Every relative rotation between parts is parameterized using the axis-angle representation. Hence, the full pose, $\theta \in \mathbb{R}^{72}$, consists of $23 \times 3 + 3$ parameters, 3 parameters per joint plus 3 for the global orientation. The global translation $\mathbf{t}$ adds 3 additional parameters. SMPL relies on a linear blend skinning (LBS) function, $W(\bar{\mathbf{V}},\mathbf{J},\theta,\mathbf{W}) : \mathbb{R}^{3N\times 3K\times|\theta|\times|\mathbf{W}|} \mapsto \mathbb{R}^{3N}$, that takes the unposed vertices in the rest pose (or zero pose), $\bar{\mathbf{V}}$, joint locations, $\mathbf{J}$, a pose, $\theta$, and the blend weights, $\mathbf{W}$, and returns the posed vertices. SMPL effectively parametrizes the skinning function with pose and shape by

$$M(\beta,\theta) = W(T_P(\beta,\theta), J(\beta), \theta, \mathbf{W}) \quad (1)$$

$$T_P(\beta,\theta) = \bar{\mathbf{T}} + B_S(\beta) + B_P(\theta) \quad (2)$$

where $B_S(\beta) \in \mathbb{R}^{3N}$ and $B_P(\theta) \in \mathbb{R}^{3N}$ are vectors of vertices representing offsets from the mean shape $\bar{\mathbf{T}}$. We name them shape and pose blend shapes respectively. The joint locations are inferred using a learned sparse regressor matrix, $\mathbf{J}_{\text{reg}} \in \mathbb{R}^{3K\times 3N}$, from the unposed shape, i.e. $J(\beta) = \mathbf{J}_{\text{reg}}(\bar{\mathbf{T}} + B_S(\beta))$. For more details please refer to [44].

#### 3.4.2 The relation of 3D Measures and volume to the SMPL shape space

One important property of SMPL is that the body vertices have a linear relation with the underlying shape space. Because the vertex shape space was learned using principal component analysis (PCA), the relation between a modification of a shape space parameter and the subsequent modification of the Euclidean distance between two vertices, is purely linear. 3D measurements relate linearly to the SMPL shape space. Note that the volume of a mesh is computed by adding the signed volumes defined by every mesh triangle and an arbitrary point. This computation has the nice property that the relationship between a shape space parameter displacement and the change in volume is cubic; i.e. each vertex undergoes a linear displacement with a modification of a shape space parameter. Thus, the cubic root of the volume of a mesh also has a linear relationship with the shape space parameters.

#### 3.4.3 Measurements to body shape regressors

The linear relation between the measurements and the body shape space is one of our key observations allowing us to train simple, yet accurate, linear regressors. We build 4 different regressors, taking respectively, 2, 4, 5 and 6 measurements as input: *R2* (from *Overall Height* and *Weight*), *R4* (from *Overall Height*, *Weight*, *Arm Span Fingers* and *Inseam Height*), *R5* (from *Overall Height*, *Weight*, *Arm Span Fingers*, *Inseam Height* and *Hips width*) and *R6* (from *Overall Height*, *Weight*, *Arm Span Fingers*, *Inseam Height*, *Hips width* and *Arm Length*).

We use the first 10 PCA components of the SMPL body shape space, and generate training subjects by sampling all the corners at the {-2, +2} standard deviations locations. Specifically we generate and measure $2^{10} = 1024$ bodies. We construct a $1024X6$ matrix containing the 6 measurements for all the 1024 bodies. From this matrix we learn the five linear regressors with a simple least squares computation.

This step completes The Virtual Caliper. Now, a user points the wand controllers of HTC Vive to the defined positions to take the measurements. These are fed to the regressors, and the SMPL model creates an accurate avatar from the resulting shape parameters.

## 4 Experimental Evaluation

The goal of the experimental evaluation is to find the best performing regressor in terms of metric accuracy, which provides a visually plausible body. The perceptual validation is important for two reasons. First, some generated bodies with reasonable metric accuracy may not look human at all. Second, it enables us to investigate to what extent the generation of metrically accurate avatars can rely on human perception [25]. In our work we metrically evaluate the created avatars in a *static* configuration. As we do not address the inverse kinematics problem, in our quantitative evaluation we do not evaluate the joint locations of the created avatars.

In this section we first present the methods we compare to (Sec. 4.1), then we evaluate metrically (Sec. 4.2) and perceptually (Sec. 4.3). We conclude discussing the metric and perceptual results (Sec. 4.4).

### 4.1 Methods for comparison

To compare metrically, we generated bodies using the open-source methods MakeHuman [5] and Unite the People [40]. In addition, we compare perceptually to BodyVisualizer [3].

*Make Human (MH1, MH2). MakeHuman* is an open-source toolkit that can be used to create a personalized avatar. Specifically, one

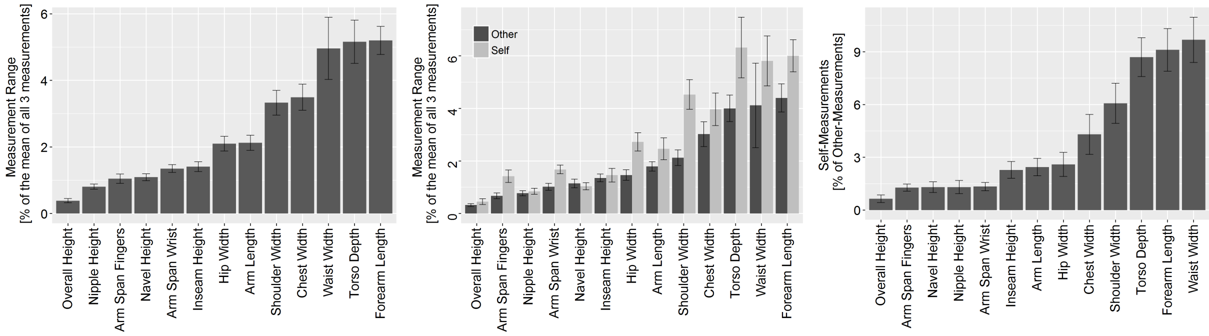


Fig. 5. Left: Repeatability of the measurements. Middle: Repeatability values for the measurer condition (*self* - *other*). Right: Accuracy of the measurements.

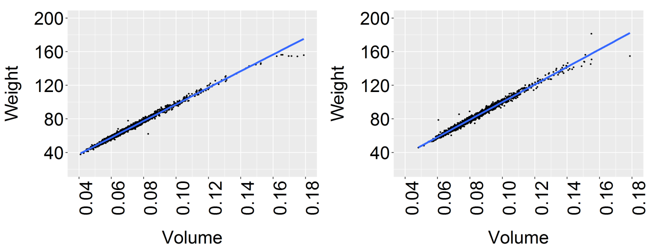


Fig. 6. Relation between the volume and the weight for the CAESAR dataset subjects. We computed the volume by registering the SMPL model to the CAESAR scans. The weight of each subject is provided in the CAESAR measurements dataset. Volume and weight are heavily linearly correlated. The linearity seems to break down for the very heavy subjects. Left: female subjects. Right: male subjects.

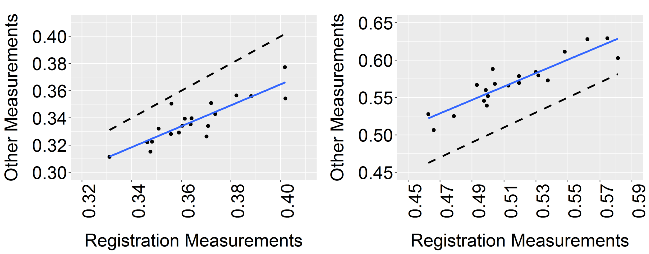


Fig. 7. Relation between the 3D measurements on the SMPL mesh alignment to the subjects scans and the 3D measurement performed by the *other* measurer. Left: *Hips Width* values. Right: *Arm Length* values. The dashed line indicates the identity line for scale clarity.

can set the height (in meters) and the weight (as a percentage of the average) of the avatar. Under the "measures" tab one can additionally set the dimensions of many local body parts in terms of dimensions and circumferences. Importantly, while adjusting height and weight influences the whole body, the changes to specific body part dimensions do not influence the measurements of neighboring body parts. For each participant we created two *MakeHuman* avatars. *MH1* is a height and weight-matched avatar, where 100% in *MakeHuman* was set to be the world average in BMI (female = 25, male = 26), and a participant's BMI was calculated as a percentage of this world average. *MH2* was matched in height and weight, and additionally the arm length (upper arm, and lower arm length) as well as inseam height (upper leg, and lower leg length) were adjusted to match the participant's measurements.

*Unite the People (UPC, UPM).* We used two pictures of the participants, in *minimal* and *cloth* conditions and obtained one body shape per picture using Lassner et al. [40] (see Fig. 8). We refer to the cloth condition result as *UPC*, and to the minimal condition as *UPM*. It is important to note that Unite the People uses a gender-neutral body model. The authors informed us that the method could not easily be adapted to gender as the CNN was trained using a gender-neutral model. While this allows the method to be fully automatic, the resulting body shapes for both males and females look somewhat male.

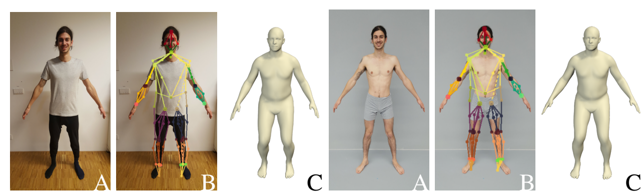


Fig. 8. Results obtained with *Unite the People*. First triplet in clothed condition (*UPC*). Second triplet in minimal condition (*UPM*). A: input image. B: detected body parts. C: obtained body shape and pose.

*BodyVisualizer (BV).* BodyVisualizer [3] is an web-based tool where one enters anthropometric measurements of a person and visualizes the predicted body. As the web application does not export the result, we could not compare to BodyVisualizer metrically. However, for the perceptual evaluation, for each participant, we used BodyVisualizer to create a body using body measurements performed by an ISAK-certified measurer (see Sec. 3.1.3) and a screen-shot of the resulting body was taken. We refer to this body shape as *BV*.

*Body Scan (Fit, Scan).* We used the 3D scan of the participants (see Sec. 3.1.3) to compute two body shape meshes. The public SMPL model approximates the human body using 10 principal components. Consequently we evaluate how well our predicted avatars match the SMPL approximation of a subject. To that end, we optimize for the first ten components of the shape space, $\beta$, that best fit the scan data in terms of point-to-surface distance. We refer to the obtained mesh as the *Fit*. In addition, we allow the SMPL mesh surface to freely deform in order to best fit the scan data in terms of point-to-surface distance. We consider the obtained mesh to be the reference and refer to it as *Scan*.

*Certified Anthropometric Measurements.* To compare the 3D measurements to a reference, an ISAK certified technician took 3D measurements using measuring tape and mechanical calipers.

### 4.2 Quantitative evaluation

We quantitatively assess four aspects of our method: 1) the regressors' precision by comparing the expected and obtained values; 2) the measurements obtained with The Virtual Caliper by comparing participants' 3D measurements to anthropometric measurements; 3) the difference between the obtained mesh surfaces to the body surfaces (Fit and Scan) obtained with the body scanner; and 4) the robustness of The Virtual Caliper under different clothing conditions.

Regressor Precision. A regressor takes a set of measurements as input, and creates a body that should ideally fulfill these measurements. For each regressor (*R2*, *R4*, *R5*, and *R6*) we used the acquired measurements (*self* and *other*) and computed the mean absolute error between

Fig. 9. Unnatural body shapes generated with the R6 regressor and the self measurements.

| | H | W (kg) | AS | I | HW | AL |
|---|---|---|---|---|---|---|
| R2 Mean | 0.00 | 0.11 | **4.26** | **2.53** | **1.48** | **1.43** |
| R2 Max | 0.00 | 0.51 | **9.76** | **5.29** | **7.29** | **3.41** |
| R4 Mean | 0.00 | 0.13 | 0.00 | 0.00 | **1.78** | **1.49** |
| R4 Max | 0.00 | 0.59 | 0.00 | 0.00 | **6.90** | **5.40** |
| R5 Mean | 0.00 | 0.19 | 0.00 | 0.00 | 0.11 | **1.50** |
| R5 Max | 0.00 | 1.38 | 0.00 | 0.00 | 0.26 | **4.72** |
| R6 Mean | 0.00 | 0.61 | 0.00 | 0.00 | 0.12 | 0.00 |
| R6 Max | 0.00 | 5.32 | 0.00 | 0.00 | 0.24 | 0.11 |
| R6o Mean | 0.00 | 0.32 | 0.00 | 0.00 | 0.11 | 0.00 |
| R6o Max | 0.00 | 0.85 | 0.00 | 0.00 | 0.18 | 0.00 |

Table 1. Metric differences between the input measurements obtained with the *The Virtual Caliper* and the resulting avatar for the different regressors, reported as the mean absolute error of all generated body shapes (self and other measurements). Units are in *cm* for all values but weight (W), in *kg*. Values in bold were not used as input in the respective regressor. Legend: H: Overall Height, W: Weight, AS: Arm Span Fingers, I: Inseam Height, HW: Hip Width, AL: Arm Length.

the input measurements and the measurements of the produced avatar. Table 1 reports the mean absolute error and max error averaged across *self* and *other* measurements for all subjects. The values not used in the regressor are shown in bold. As expected, due to the linear relationship between the measurements and the shape space parameters, the errors between the input and output body dimensions are consistently zero or very small; i.e. on the order of millimeters for the *Hip Width* and *Arm Length*. Minor errors for the weight occur for *R2*, *R4*, and *R5*, (averaging 200 grams). These errors are likely to arise due to rounding errors in the computations. *R6* showed a significantly higher maximum error of 5 kg. When visually examining the bodies, *R6* was found to produce unnatural body shapes (see Fig. 9). Therefore, we also computed the error using only the measurement values of the *other* measurer, resulting in a significantly reduced error. We refer to the versions using self and other measurements as *R6s* and *R6o* respectively and use this notation in the remainder of the paper.

HTC-Vive self and other to anthropometric measurements. To investigate the precision of the measurements using The Virtual Caliper (for *self* and *other* measurers), we compared the measurements to the anthropometric hand measurements performed by the ISAK certified measurer. The Mean Absolute Errors (MAE) and Relative Errors (RE) between the measurements are shown in Table 2. Most errors are below 2% both for *self* and *other* measurements. The biggest error for *self* measurements was observed for the *Hip Width* and *Arm Length* indicating that those are either more difficult to measure, or instructions about the measurement location should be improved.

Body surface distance. The surface-to-surface distance between the generated bodies and the *Scan* was computed. This is especially important as the measurements of the generated bodies could be precise even when the generated body has a non-human-like body shape, as we observed for *R6s*. As the surface distance between two meshes with different heights depends on their rigid registration relative to each other, all avatars were set on a common floor and centered horizontally and in depth using their bounding boxes.

For this evaluation we considered an additional variant of the R5 regressor: *R5o* and *R5s* which use the raw values of the HTC Vive, and *R5sL* and *R5oL* which use the learned mapping in Section 3.3 for the *Hips Width* measurement. The goal is to quantitatively evaluate the impact of the learned mapping on the surface distances. With a consistent notation we refer with *R6sL* and *R6oL* to the *R6* regressor using respectively the *self* and *other* measurements, and with the learned mappings from Section 3.3.

Table 4 reports the numeric surface distance values, and Figure 10 shows the heat maps of the errors on the average female body. The results of the computed surface-to-surface distances show that the proposed regressors metrically outperform *MakeHuman* and *Unite the People* in terms of their deviation from participants' body scans. R2o, R4o and R5oL obtain the best metric results with an accuracy of 1.11-1.18 cm. The results also show that for the proposed methods, the *other* measurements consistently have less error than *self* measurements. It is interesting to note that because we computed the avatars with the *R5* regressors without the *Hips Width* correction, the lower torso part is excessively protruded. *R5sL* and *R5oL*, using the *Hips Width* mapping, properly recover this error and better explain the hip area.

Robustness to clothing conditions. To evaluate the robustness of our method with respect to different clothing conditions, we invited 8 participants back (gender balanced), and asked them to perform the measurements in three types of clothing: *minimal*, *loose* (wearing long sports pants, and a loose t-shirt), and *own* clothes (the street clothes they were naturally wearing). The results are presented in Table 3. The relative errors in the measurements are consistent with the ones observed in Table 2 and the relative ranges consistent with the observed ranges in Figure 5. The Virtual Caliper is robust to different clothing conditions and does not require minimal clothing to obtain faithful measurements.

| 3D Meas. | MAE *self* | MAE *other* | RE *self* | RE *other* |
|---|---|---|---|---|
| Overall Height | 1.53 cm | 1.47 cm | 0.87% | 0.85% |
| Arm Span | 2.01 cm | 1.72 cm | 1.16% | 1.02% |
| Arm Length | 1.29 cm | 0.76 cm | 2.27% | 1.34% |
| Inseam Height | 2.06 cm | 2.46 cm | 1.50% | 1.84% |
| Hip Width | 0.77 cm | 0.65 cm | 2.35% | 1.93% |

Table 2. Mean Absolute Errors (MAE) and Relative Error (RE) between the measurements obtained with the *The Virtual Caliper* for the *self* and *other* measurer, compared to ISAK certified measurements.

| | RE | | RR | | |
|---|---|---|---|---|---|
| | *own* | *loose* | *minimal* | *own* | *loose* |
| Overall Height | 0.41% | 0.50% | 0.21% | 0.28% | 0.27% |
| Arm Span | 0.81% | 0.70% | 0.98 % | 1.10% | 0.66% |
| Arm Length | 2.01% | 1.23% | 2.63% | 3.35% | 3.03% |
| Inseam Height | 1.6% | 1.24% | 1.60% | 1.77% | 0.96% |
| Hip Width | 0.84% | 1.51% | 1.59 % | 3.20 % | 1.59% |

Table 3. Relative Error (RE) between measurements by the same subjects in different clothes. Clothing *minimal* is used as reference. Relative Range (RR) between measurements obtained with different clothes.

### 4.3 Perceptual Evaluation

To examine how the bodies generated with the different techniques are perceived and evaluated, all participants from the first study (see Sec. 3.1.3) were invited back for a perceptual evaluation session comprising three short experiments (see Fig. 11). Out of the 20 participants who were scanned and measured, 18 (nine males, nine females) took part in the perceptual study. The perceptual evaluation session took place five weeks after the scan session and took approximately 60 minutes.

In two of the three experiments, participants evaluated images of personalized bodies in terms of similarity to their own body. For each participant, 15 different body stimuli were generated and images of the body stimuli were rendered. As in the quantitative evaluation, the stimuli set comprised the *Fit*, the bodies generated based on the *self* and *other* measurements using *R2*, *R4*, and *R5*, as well as *R5oL*,

| | | proposed | | | | | | | | | | previous | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | fit | R2s | R2o | R4s | R4o | R5s | R5o | R5sL | R5oL | R6sL | R6oL | MH1 | MH2 | UPM | UPC |
| Mean | 0.72 | 1.25 | 1.12 | 1.37 | 1.11 | 1.59 | 1.28 | 1.44 | 1.18 | 2.05 | 1.43 | 3.38 | 3.57 | 5.66 | 7.84 |
| Max | 3.24 | 4.64 | 4.17 | 4.86 | 4.17 | 5.58 | 4.62 | 5.10 | 4.28 | 7.09 | 4.85 | 12.4 | 12.5 | 16.9 | 21.8 |

Table 4. Surface to surface results of the created avatars. Mean and Max reported values are in centimeters.

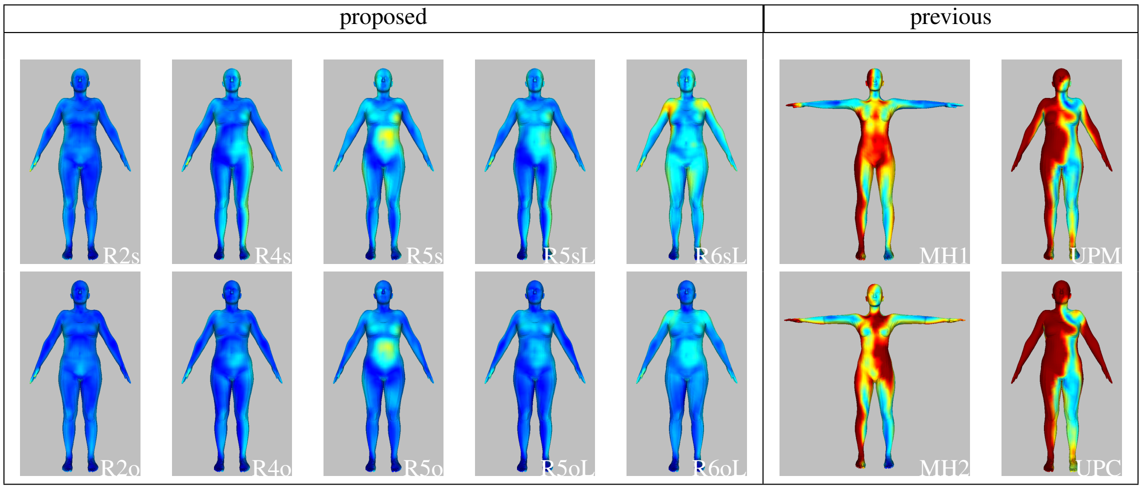


Fig. 10. Visualization of the surface errors displayed on the average female body for our proposed and previously available methods. Dark blue indicates zero error, red indicates an error of 5 cm or more. The proposed methods outperform the previous methods in terms of error when compared to the scan. In *R5L*, hip width correction is applied resulting in a reduced error when compared to *R5* both in self and other condition. Notice that *R6sL* resulting in unnatural body shapes has a high surface error.

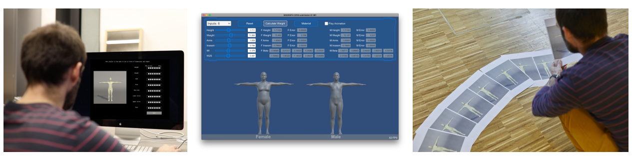

Fig. 11. Perceptual evaluation. Participants performed two tasks on the desktop and ranked printed images.

*R6oL*, *MakeHuman* (*MH1*, *MH2*), and *United the People* (*UPC*, *UPM*). Additionally, a body generated with *BodyVisualizer* was added, as well as the *Scan*. We predicted that the scan would receive the highest similarity ratings as it contains most identity-specific features. Further, as the *Fit* is a good approximation of the body shapes, it should also be among the highest rated bodies.

In the third experiment, participants adjusted an avatar to their perceived own body dimensions (method of adjustment) using three of our developed regressors. Participants completed the experiments in the same order as presented below.

#### 4.3.1 Experiments

**Overall Similarity Ratings.** The experiment was created in OpenSesame (v.3.1.9) [46] and presented on a 21″monitor. In each trial, participants were presented with one of the 15 personalized images and were asked to rate how similar the body was to their own body on a 7-point Likert scale (1 - not at all, 7 - very). Participants could view each image as closely and as long as they wanted. The order of the image presentation was randomized across participants. The task took approximately 5 minutes.

**Ranking.** For each participant, the 15 personalized bodies were printed on A4 paper sheets. Participants stood at a large table and were instructed to carefully view the bodies and arrange them from least (left) to most (right) similar to their own body in terms of body dimensions and shape. Once completed they wrote the ranking on the images from 1 to 15, resulting in a forced-ranking of all bodies. The task took approximately 5-10 minutes.

**Desktop Method of Adjustment.** For the method of adjustment task, 3 different regressors were used: *R2*, *R4*, *R6* where the body dimensions could be individually adjusted. The desktop application was programmed in Unity game engine and was displayed on a 21″monitor. Since many on-line tools for clothing try-on allow the user to enter the measurement values directly, participants were asked to adjust a gender-matched avatar to their own body dimensions with the adjusted metric values being visible. They were asked to perform this task hierarchically for three regressors, by first adjusting height and weight, then inseam height and arm span, and finally arm length and hip width. Participants were specifically instructed to avoid creating statistically implausible body shapes as indicated by the shape space values shown in the application turning red. Once the participants were done adjusting the body, the six values were recorded. The task took approximately 10 minutes.

#### 4.3.2 Results

**Overall Similarity Ratings.** Figure 12 shows the results of the overall similarity ratings. As expected, the *Scan* was rated as most similar to the participants' bodies. To examine which of the methods received significantly lower similarity ratings as compared to the *Scan*, planned comparisons using paired t-tests with p-value correction for multiple comparisons were conducted. The results show that MH2, UPC, and UPM received lower similarity ratings than the *Scan* (all p-values $< .05$). There was no significant difference between the similarity ratings of the *Scan* and BV, R2s, Fit, R5oL, R4s, R2o, R4o, R5o, R5s, R6oL, and MH1, indicating that the bodies generated with these methods were statistically similarly rated.

**Ranking.** Figure 13 shows the results of the image ranking task. To examine which of the methods received significantly lower ranking values as compared to the *Scan*, planned comparisons using paired t-tests with p-value correction for multiple comparisons were conducted. The results show that R5o, R5s, R6oL, MH1, MH2, UPC, and UPM received significantly lower ranking values (all p-values $< .05$). There was no significant difference between the ranking of the scan and the ranking of BV, R2s, Fit, R5oL, R4s, R2o, and R4o, indicating that they are statistically visually similar.

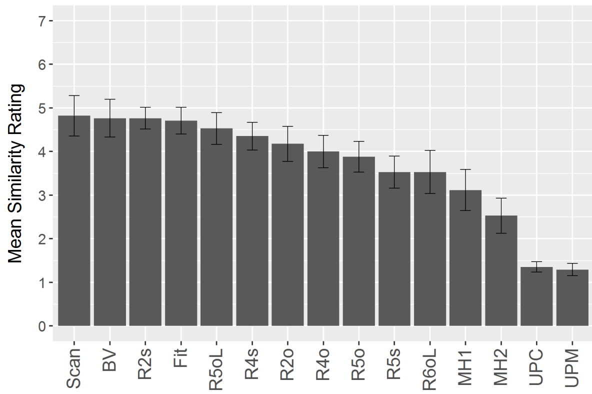


Fig. 12. Overall similarity ratings of the 15 different body images. The methods are ranked with respect to their mean rating. The error bars represent standard errors of the mean.

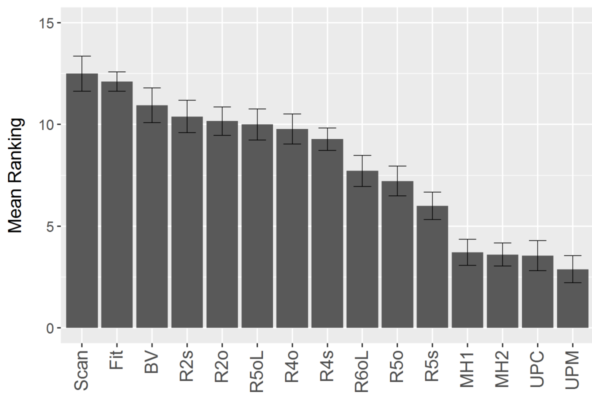


Fig. 13. Rankings of the 15 body images. Least (1) to most (15) similar.

Method of Adjustment. For each participant, the resulting body from the Method of Adjustment task was compared to the body dimensions of the *Fit*. The signed percent errors are presented in Figure 14. Since the metric values were visible, participants were accurate when adjusting their body heights. However, in terms of weight, arm span, inseam, arm lengths and hip width the adjusted values were not so accurate, with e.g. errors over 5% for the arm span and arm length. This finding is in line with previous studies on visual body perception showing that one can not rely on human perception to create metrically accurate bodies [25].

### 4.4 Discussion of Experimental Results

As stated earlier, the goal is to find the best regressor providing metric accuracy as well as a perceptually appealing body. Table 1 indicates that regressors with the most measurements should be preferred, as they faithfully generate bodies with the accurate measurements. However, in our study we reached the limits of our approach with *R6*. The *self* measurements produced unrealistic avatars (Fig. 9), and even the bodies created with R6oL were poorly ranked. The regressor *R5L*, which uses *Overall Height*, *Weight*, *Arm Span Fingers*, *Inseam Height* and *Hips Width* as input, generates accurate metric results both in the dimensions (Table 1) and in surface error (Table 4). It also obtains as good overall similarity ratings as the *Scan* (Fig. 12) and is among the best ranked bodies (Fig. 13). While *R4* obtains slightly better metric surface errors than R5L (Table 4) and good perceptual ratings (Fig. 12) and rankings (Fig. 13), it has errors of almost 7 cm in the *Hip Width* dimension (Table 1) as it is not used as input. For these reasons we conclude that the best overall performing regressor is R5L.

In a second level, we confirmed the hypothesis that human perception is not suited to create metrically accurate avatars [25]. Our perceptual

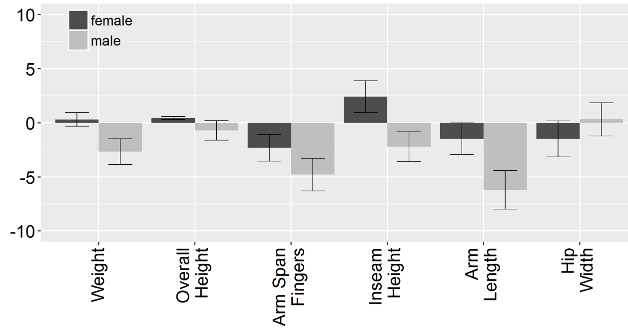


Fig. 14. Signed percent error of the adjusted body in the Method of Adjustment task on the Desktop compared to the Fit (by gender condition). Negative values indicate underestimation of body dimensions as compared to the Fit.

study provides evidence in this direction, showing that even bodies with significant differences in the *Hips Width* obtained high overall rating, for instance *R2o* and *R2s* having mean errors of 2cm. This further argues for the need of The Virtual Caliper. Users can not rely on their perception and need a tool to precisely measure themselves.

The metric evaluation results showed that our methods systematically outperform the open-source state-of-the-art methods MakeHuman and Unite the People. However, we would like to point out that the achieved metric accuracy when comparing to the *Scan* is on the order of $1.1cm$, and so The Virtual Caliper is less accurate than the RGB-D methods such as [16, 41, 74], reporting errors on the order of $2.4mm$, $3mm$ and $2.4mm$ respectively. If the protection of privacy and identity are not required, these methods should be preferred in terms of accuracy.

Another limitation of our approach is that it does not capture the face. For applications where the face is important for identifying an individual, already available hardware such as the iPhoneX could be used; the AR-Kit extracts a personalized mesh of the face, which could be combined with our mesh.

## 5 Rapid Avatar Creation Tools

Two pieces of software are made available[2]. The first tool is The Virtual Caliper; i.e. the HTC Vive measurement tool. The user is guided with simple videos to perform the 3D measurements needed for the creation of the avatars. In order to account for the offset of the HTC Vive initial calibration [51], the user is instructed to measure the floor at five locations. Then the user is guided to perform the 3D measurements.

The second tool is an interactive Desktop application integrating the regressors, R2, R4, R5 and R6. Users can manually set measurement values and export the created avatars in the FBX file format. The values from the first tool can be used in the second one to generate a rigged avatar model that matches the measured user body dimensions. Because the avatars are built upon the openly available SMPL model [44], they can be easily animated and posed in real-time. The desktop application is available on OSX and Windows.

## 6 Conclusions

We presented The Virtual Caliper, a tool that empowers users to measure their own bodies using an inexpensive and easy to use VR system (HTC Vive) and create metrically precise animation-ready avatars. We also provide a desktop application for creating bodies and editing body shape resulting in avatars that accurately match the users' measured dimensions. Our approach overcomes potential privacy issues with previous methods in that no pictures of the user are required and the protocol can be performed in clothing.

While the produced software tools use the HTC Vive, any other device precisely measuring 3D distances, such as Oculus Rift controllers or reflective markers for motion capture systems, could also be used. As we only require users to have the "basic HTC Vive configuration"

[2]www.thevirtualcaliper.is.tue.mpg.de

to perform 3D measurements, we do not address the inverse kinematics (IK) problem in this work. Although our solution does not provide an IK solver, the generated avatar is compatible with third-party commercial IK solvers like IKinema or FinalIK (Unity).

With the obtained results, several future work directions will be pursued. Our results show that the precision of the measurements play an important role, especially as the number of used measurements increases. The user performance has two tangled factors: the landmark location and the human pose. The first is the capability of the user to identify the body landmark where the measurement needs to be taken. Identifying which landmarks are not properly understood will allow us to improve the instructional videos. The second is the pose adopted by the user while performing the measurement. We intend to address this issue in future work by equipping the users with motion capture markers [43] or IMU sensors [64] in order to obtain the pose of the subject while performing the measurements. Identifying biases due to pose could also be important for improving the instructional videos.

As we deliberately decided to exclude pictures from The Virtual Caliper, the generated avatars do not include texture. If the user wants to add texture to the avatar, one could adapt the optimization in [40] and instead of optimizing for shape and pose, one could use the obtained shape from The Virtual Caliper and just optimize for pose. With the image and mesh geometry registered, the image information could be used to texture the avatar.

The Virtual Caliper does not include any measurement of the waist and chest regions as the combination of soft tissue as well as the breathing cycle makes it difficult to obtain accurate measurements. These body regions are known to be challenging even for anthropometric measurements [52]. We envision several ways to overcome this limitation in the future. Similar to the pipeline proposed by Wuhrer et al. [68], the user could generate an initial body using the current version of The Virtual Caliper and locally refine the body in an AR or VR environment.

So far we only considered punctual measurements. The user holds the trigger of the wand controller for one second and the measurement is saved. With the gathered experience we will further explore which continuous measurements could be relevant to further refine the body shape. For example, dynamic measurements of the torso during the breathing cycle. So far, anthropometric measurement procedures have never had the chance to accurately assess these dynamics.

## Disclosure

All of the work for this paper performed by Betty Mohler was done prior to her employment with Amazon. Michael J. Black has received research gift funds from Intel, NVIDIA, Adobe, Facebook, and Amazon. While MJB is a part-time employee of Amazon, this research was performed solely at MPI. Naureen Mahmood is a founder and shareholder of Meshcapade GmbH, which is commercializing body shape technology. All research for this project was performed solely at, and funded solely by, MPI.

## Acknowledgments

Authors wish to acknowledge Andrea Keller and Tsvetelina Alexiadis for helping with the data collection as well as Thomas Motta and Jan Rottler for the footage acquisition and processing. In loving memory of Daniel Cudeiro (1995-2018).

## References

[1] [Tutorial] Calibrating Height and Arm Length in VR with Unity3d. https://www.youtube.com/watch?time_continue=2&v=-5GvMk4zRWs, 2017. Accessed: 2018-06-01.

[2] AXYZdesign. https://secure.axyz-design.com/metropoly-3d-people, 2018. Accessed: 2018-01-18.

[3] BodyVisualizer. http://www.bodyvisualizer.com/, 2018. Accessed: 2018-01-18.

[4] International Society of the Advancement of Kinanthropometry: ISAK. http://www.isak.global/, 2018. Accessed: 2018-01-18.

[5] MakeHuman. http://www.makehuman.org/, 2018. Accessed: 2018-01-18.

[6] Mixamo. https://www.mixamo.com/, 2018. Accessed: 2018-01-18.

[7] Poser 3D Character Art and Animation Software. http://my.smithmicro.com/poser-3d-animation-software.html, 2018. Accessed: 2018-01-18.

[8] J. Achenbach, T. Waltemate, M. Latoschik, and M. Botsch. Fast generation of realistic virtual humans. In *Proceedings of ACM Symposium on Virtual Reality Software and Technology*, 2017.

[9] B. Allen, B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parameterization from range scans. In *ACM transactions on graphics (TOG)*, vol. 22, pp. 587–594. ACM, 2003.

[10] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. Scape: shape completion and animation of people. In *ACM Transactions on Graphics (TOG)*, vol. 24, pp. 408–416. ACM, 2005.

[11] N. Badler. Virtual humans for animation, ergonomics, and simulation. In *Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE*, pp. 28–36. IEEE, 1997.

[12] A. O. Balan, L. Sigal, M. J. Black, J. E. Davis, and H. W. Haussecker. Detailed human shape and pose from images. In *Conference on Computer Vision and Pattern Recognition*, pp. 1–8. IEEE, 2007.

[13] A. Bauer, A.-H. Dicko, F. Faure, O. Palombi, and J. Troccaz. Anatomical mirroring: Real-time user-specific anatomy in motion using a commodity depth camera. In *Proceedings of the 9th International Conference on Motion in Games*, MIG '16, pp. 113–122. ACM, New York, NY, USA, 2016. doi: 10.1145/2994258.2994259

[14] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 187–194. ACM Press/Addison-Wesley Publishing Co., 1999.

[15] K. J. Blom, A. I. B. Rivas, X. Alvarez, O. Cetinaslan, B. Oliveira, V. Orvalho, and M. Slater. Achieving participant acceptance of their avatars. *Presence: Teleoperators and Virtual Environments*, 23(3):287–299, 2014. doi: 10.1162/PRES_a_00194

[16] F. Bogo, M. J. Black, M. Loper, and J. Romero. Detailed full-body reconstructions of moving people from monocular RGB-D sequences. In *International Conference on Computer Vision*, pp. 2300–2308, Dec. 2015.

[17] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. J. Black. Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. In *European Conference on Computer Vision*, Lecture Notes in Computer Science. Springer International Publishing, Oct. 2016.

[18] J. Boisvert, C. Shu, S. Wuhrer, and P. Xi. Three-dimensional human shape inference from silhouettes: reconstruction and validation. *Machine vision and applications*, 24(1):145–157, 2013.

[19] C.-H. Chu, Y.-T. Tsai, C. C. Wang, and T.-H. Kwok. Exemplar-based statistical model for semantic parametric design of human body. *Computers in Industry*, 61(6):541–549, 2010.

[20] F. Cordier, H. Seo, and N. Magnenat-Thalmann. Made-to-measure technologies for an online clothing store. *IEEE Computer graphics and applications*, 23(1):38–48, 2003.

[21] D. DeCarlo, D. Metaxas, and M. Stone. An anthropometric face model using variational techniques. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pp. 67–74. ACM, 1998.

[22] E. Dibra, C. Öztireli, R. Ziegler, and M. Gross. Shape from selfies: Human body shape estimation using cca regression forests. In *European Conference on Computer Vision*, pp. 88–104. Springer, 2016.

[23] O. Freifeld and M. J. Black. Lie bodies: A manifold representation of 3d human shape. In *European Conference on Computer Vision*, pp. 1–14. Springer, 2012.

[24] H. Fuchs, A. State, and J.-C. Bazin. Immersive 3d telepresence. *Computer*, 47(7):46–52, 2014.

[25] C. T. Fuentes, M. R. Longo, and P. Haggard. Body image distortions in healthy adults. *Acta psychologica*, 144(2):344–351, 2013.

[26] C. C. Gordon, T. Churchill, C. E. Clauser, B. Bradtmiller, J. T. McConville, I. Tebbetts, and R. A. Walker. Anthropometric survey of us army personnel: Summary statistics, interim report for 1988. Technical report, ANTHROPOLOGY RESEARCH PROJECT INC YELLOW SPRINGS OH, 1989.

[27] P. Guan, A. Weiss, A. O. Balan, and M. J. Black. Estimating human shape and pose from a single image. In *Computer Vision, International Conference on*, pp. 1381–1388. IEEE, 2009.

[28] H. Han, Y. Nam, and K. Choi. Comparative analysis of 3d body scan measurements and manual measurements of size korea adult females. *International Journal of Industrial Ergonomics*, 40(5):530–540, 2010.

[29] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen, and H.-P. Seidel. Multilinear pose and body shape estimation of dressed subjects from

image sets. In *Conference on Computer Vision and Pattern Recognition*, pp. 1823–1830. IEEE, 2010.

[30] N. Hasler, C. Stoll, B. Rosenhahn, T. Thormählen, and H.-P. Seidel. Estimating body shape of dressed humans. *Computers & Graphics*, 33(3):211–216, 2009.

[31] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel. A statistical model of human pose and body shape. In *Computer Graphics Forum*, vol. 28, pp. 337–346. Wiley Online Library, 2009.

[32] A. Hilton, D. Beresford, T. Gentils, R. Smith, and W. Sun. Virtual people: Capturing human models to populate virtual worlds. In *Computer Animation, 1999. Proceedings*, pp. 174–185. IEEE, 1999.

[33] H. Honglun, S. Shouqian, and P. Yunhe. Research on virtual human in ergonomic simulation. *Computers & Industrial Engineering*, 53(2):350–356, 2007.

[34] A. Jain, T. Thormählen, H.-P. Seidel, and C. Theobalt. Moviereshape: Tracking and reshaping of humans in videos. In *ACM Transactions on Graphics (TOG)*, vol. 29, p. 148. ACM, 2010.

[35] P. Kadleček, A.-E. Ichim, T. Liu, J. Křivánek, and L. Kavan. Reconstructing personalized anatomical models for physics-based body animation. *ACM Transactions on Graphics (TOG)*, 35(6):213, 2016.

[36] A. Keizer, A. van Elburg, R. Helms, and H. C. Dijkerman. A virtual reality full body illusion improves body image disturbance in anorexia nervosa. *PloS one*, 11(10):e0163921, 2016.

[37] K. Kristensen, N. Borum, L. G. Christensen, H. W. Jepsen, J. Lam, A. L. Brooks, and E. P. Brooks. Towards a next generation universally accessible online shopping-for-apparelsystem. In *International Conference on Human-Computer Interaction*, pp. 418–427. Springer, 2013.

[38] H. J. Krzywicki and K. S. Chinn. Human body density and fat of an adult male population as measured by water displacement. *The American journal of clinical nutrition*, 20(4):305–310, 1967.

[39] C. Kuster, N. Ranieri, H. Zimmer, J.-C. Bazin, C. Sun, T. Popa, M. Gross, et al. Towards next generation 3d teleconferencing systems. In *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012*, pp. 1–4. IEEE, 2012.

[40] C. Lassner, J. Romero, M. Kiefel, F. Bogo, M. J. Black, and P. V. Gehler. Unite the people: Closing the loop between 3d and 2d human representations. In *Conference on Computer Vision and Pattern Recognition*, July 2017.

[41] H. Li, E. Vouga, A. Gudym, L. Luo, J. T. Barron, and G. Gusev. 3D self-portraits. *ACM Trans. Graph.*, 32(6):187:1–187:9, Nov. 2013. doi: 10.1145/2508363.2508407

[42] T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero. Learning a model of facial shape and expression from 4D scans. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 36(6), 2017.

[43] M. Loper, N. Mahmood, and M. J. Black. Mosh: Motion and shape capture from sparse markers. *ACM Transactions on Graphics (TOG)*, 33(6):220, 2014.

[44] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. Smpl: A skinned multi-person linear model. *ACM Transactions on Graphics (TOG)*, 34(6):248:1–248:16, Oct. 2015. doi: 10.1145/2816795.2818013

[45] C. Malleson, M. Kosek, M. Klaudiny, I. Huerta, J.-C. Bazin, A. Sorkine-Hornung, M. Mine, and K. Mitchell. Rapid one-shot acquisition of dynamic VR avatars. In *Virtual Reality (VR)*, pp. 131–140. IEEE, 2017.

[46] S. Mathôt, D. Schreij, and J. Theeuwes. Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, 44(2):314–324, 2012.

[47] S. Mölbert, A. Thaler, B. Mohler, S. Streuber, J. Romero, M. Black, S. Zipfel, H.-O. Karnath, and K. Giel. Assessing body image in anorexia nervosa using biometric self-avatars in virtual reality: Attitudinal components rather than visual body size estimation are distorted. *Psychological medicine*, pp. 1–12, 2017.

[48] M. Mori, K. MacDorman, and N. Kageki. The Uncanny Valley [from the field]. *IEEE Robotics Automation Magazine*, 19(2):98–100, 2012. doi: 10.1109/MRA.2012.2192811

[49] A. Neophytou and A. Hilton. Shape and pose space deformation for subject specific animation. In *3DTV-Conference, 2013 International Conference on*, pp. 334–341. IEEE, 2013.

[50] A. Neophytou and A. Hilton. A layered model of human body and garment deformation. In *2014 2nd International Conference on 3D Vision*, vol. 1, pp. 171–178. IEEE, 2014.

[51] D. C. Niehorster, L. Li, and M. Lappe. The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research. *i-Perception*, 8(3):2041669517708205, 2017.

[52] S. Paquette, C. Gordon, and B. Bradtmiller. Anthropometric survey (ansur) ii pilot study: methods and summary statistics. Technical report, ANTHROTECH YELLOW SPRINGS OH, 2009.

[53] H. Rhodin, N. Robertini, D. Casas, C. Richardt, H.-P. Seidel, and C. Theobalt. General automatic human shape and motion capture using volumetric contour cues. In *European Conference on Computer Vision*, pp. 509–526. Springer, 2016.

[54] K. M. Robinette, S. Blackwell, H. Daanen, M. Boehmer, and S. Fleming. Civilian american and european surface anthropometry resource (CAESAR), final report. volume 1. summary. Technical report, SYTRONICS INC DAYTON OH, 2002.

[55] K. M. Robinette, S. Blackwell, H. Daanen, M. Boehmer, and S. Fleming. Civilian american and european surface anthropometry resource (caesar), final report. volume 1. summary. Technical report, SYTRONICS INC DAYTON OH, 2002.

[56] K. M. Robinette and H. A. Daanen. Precision of the caesar scan-extracted measurements. *Applied Ergonomics*, 37(3):259–265, 2006.

[57] C. Rupprecht, O. Pauly, C. Theobalt, and S. Ilic. 3d semantic parameterization for human shape modeling: Application to 3d animation. In *3DTV, International Conference on*, pp. 255–262. IEEE, 2013.

[58] S. Saito, Z.-Y. Zhou, and L. Kavan. Computational bodybuilding: Anatomically-based modeling of human bodies. *ACM Trans. Graph.*, 34(4):41:1–41:12, July 2015. doi: 10.1145/2766957

[59] H. Seo and N. Magnenat-Thalmann. An example-based approach to human body manipulation. *Graphical Models*, 66(1):1–23, 2004.

[60] A. Shapiro, A. Feng, R. Wang, H. Li, M. Bolas, G. Medioni, and E. Suma. Rapid avatar capture and simulation using commodity depth sensors. *Computer Animation and Virtual Worlds*, 25(3-4):201–211, 2014.

[61] C. Stoll, J. Gall, E. de Aguiar, S. Thrun, and C. Theobalt. Video-based reconstruction of animatable human characters. *ACM Transactions on Graphics (TOG)*, 29(6):139:1–139:10, Dec. 2010.

[62] S. Streuber, M. A. Quiros-Ramirez, M. Q. Hill, C. A. Hahn, S. Zuffi, A. O'Toole, and M. J. Black. Body talk: Crowdshaping realistic 3d avatars with words. *ACM Transactions on Graphics (TOG)*, 35(4):54, 2016.

[63] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan. Scanning 3d full human bodies using kinects. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):643–650, April 2012. doi: 10.1109/TVCG.2012.56

[64] T. von Marcard, B. Rosenhahn, M. J. Black, and G. Pons-Moll. Sparse inertial poser: Automatic 3d human pose estimation from sparse imus. In *Computer Graphics Forum*, vol. 36, pp. 349–360. Wiley Online Library, 2017.

[65] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE transactions on visualization and computer graphics*, 24(4):1643–1652, 2018.

[66] A. Weiss, D. Hirshberg, and M. J. Black. Home 3d body scans from noisy image and range data. In *Computer Vision, IEEE International Conference on*, pp. 1951–1958. IEEE, 2011.

[67] S. Wuhrer, L. Pishchulin, A. Brunton, C. Shu, and J. Lang. Estimation of human body shape and posture under clothing. *Computer Vision and Image Understanding*, 127:31–42, 2014.

[68] S. Wuhrer and C. Shu. Estimating 3d human shapes from measurements. *Machine vision and applications*, 24(6):1133–1147, 2013.

[69] J. Yang, J.-S. Franco, F. Hétroy-Wheeler, and S. Wuhrer. Estimation of Human Body Shape in Motion with Wide Clothing. In *European Conference on Computer Vision*. Amsterdam, Netherlands, Oct. 2016.

[70] M. Ye and R. Yang. Real-time simultaneous pose and shape estimation for articulated objects using a single depth camera. In *Conference on Computer Vision and Pattern Recognition*, pp. 2345–2352. IEEE, 2014.

[71] C. Zhang, S. Pujades, M. J. Black, and G. Pons-Moll. Detailed, accurate, human shape estimation from clothed 3d scan sequences. In *Conference on Computer Vision and Pattern Recognition*, vol. 00, pp. 5484–5493, July 2017. doi: 10.1109/CVPR.2017.582

[72] Q. Zhang, B. Fu, M. Ye, and R. Yang. Quality dynamic human body modeling using a single low-cost depth camera. In *Conference on Computer Vision and Pattern Recognition*, pp. 676–683. IEEE, 2014.

[73] S. Zhou, H. Fu, L. Liu, D. Cohen-Or, and X. Han. Parametric reshaping of human bodies in images. In *ACM Transactions on Graphics (TOG)*, vol. 29, p. 126. ACM, 2010.

[74] M. Zollhöfer, M. Niessner, S. Izadi, C. Rehmann, C. Zach, M. Fisher, C. Wu, A. Fitzgibbon, C. Loop, C. Theobalt, and M. Stamminger. Real-time Non-rigid Reconstruction Using an RGB-D Camera. *ACM Trans. Graph.*, 33(4):156:1–156:12, July 2014. doi: 10.1145/2601097.2601165